

#4

Docket No. 34645-00516USPX

P11323US1

3490 U.S. PAT. 09/05/00 09/05/00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Jim Sundqvist et al.) Group Art Unit: not assigned
Serial No.: not assigned) Examiner: not assigned
Filed: herewith)
For: METHOD AND APPARATUS IN A TELECOMMUNICATIONS SYSTEM

Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

CERTIFICATE OF MAILING BY EXPRESS MAIL

"EXPRESS MAIL" Mailing Label No EL524957506US

Date of Deposit: Sept 5, 2000
I hereby certify that this paper or fee is being deposited with the U.S. Postal Service
"Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated
above and is addressed to the Assistant Commissioner for Patents, Washington, D.C.
20231

Type or Print Name

DOROTHY MACKINNON

Signature

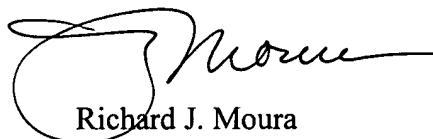
Dorothy MacKinnon

Dear Sir:

CLAIM OF PRIORITY UNDER 35 U.S.C. § 119

Under the provisions of 35 U.S.C. 119 Applicant hereby claims the priority of Swedish patent application no. 9903223-7 filed on September 9, 1999, which is mentioned in the declaration of the above-identified application. A certified copy of the priority document is filed herewith.

Respectfully submitted,



Richard J. Moura
Reg. No. 34,883

Jenkins & Gilchrist, P.C.
1445 Ross Avenue, Suite 3200
Dallas, Texas 75202-2799
214/855-4709 (Direct)
214/855-4300 (Fax)

PRV

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

Intyg Certificate

Jc490 U.S. PTO
09/655326
109/05/00



Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

(71) Sökande Telefonaktiebolaget L M Ericsson, Stockholm SE
Applicant (s)

(21) Patentansökningsnummer 9903223-7
Patent application number

(86) Ingivningsdatum 1999-09-09
Date of filing

Stockholm, 2000-07-04

För Patent- och registreringsverket
For the Patent- and Registration Office

A. Södervall
Anita Södervall

Avgift
Fee 170.-

**PATENT- OCH
REGISTRERINGSVERKET
SWEDEN**

Postadress/Adress
Box 5055
S-102 42 STOCKHOLM

Telefon/Phone
+46 8 782 25 00
Vx 08-782 25 00

Telex
17978
PATOREG S

Telefax
+46 8 666 02 86
08-666 02 86

METHOD AND APPARATUS IN A TELECOMMUNICATIONS SYSTEM**TECHNICAL FIELD OF THE INVENTION**

The present invention relates generally to methods for improving speech quality in e.g. IP-telephony systems. More particularly the present invention relates to a method for
5 reducing audio artefacts due to overrun or underrun in a playout buffer.

The invention also relates to an arrangement for carrying out the method.

10 DESCRIPTION OF RELATED ART

When sampling frequencies, in e.g. a speech coding system, are not controlled, underrun or overrun might occur in the playout buffer, which is a buffer storing speech samples for later
15 playout. Underrun means that the playout buffer will run into starvation, i.e. it will no longer have any samples to play on the output. Overrun means that the playout buffer will be filled with samples and that following samples cannot be buffered and consequently will be lost. Underrun is probably
20 more common than overrun since the size of the playout buffer can increase until there is no memory left, while it only can decrease until there are no samples left.

Currently, most systems do not deal with the problem that the sampling frequency might differ considerably between the sending and the receiving side. One possible solution proposed
25 in, EP-0680033 A2, works on pitch periods. Adding or removing pitch periods in the speech signal achieves a different duration of a speech segment without affecting other speech characteristics than speed. This proposed solution might be used as an indirect sample rate conversion method.

Another solution uses the beginning of talkspurts as an indication to reset the playout buffer to a specified level. The distance, in number of samples, between two consecutive talkspurts is increased if the receiving side is playing faster than the sending side and decreased if the receiving side is playing slower than the sending side. In IP-telephony solutions, using the IP/UDP/RTP-protocols (Internet Protocol/User Datagram Protocol/Real Time Protocol); the marker flag in the RTP header is used to identify the beginning of a talkspurt. At the beginning of a talkspurt the playout buffer is set to a suitable size.

The solution according to EP-0680033 A2, where pitch periods are removed or inserted, assumes a fixed conversion factor between the receiving and transmitting side. Therefore it cannot be used in dynamical systems, i.e. where the sampling frequencies varies. Further, it does not solve the problem with underrun or overrun situations, but is instead focused on changing the playback rate of a speech signal stored in compressed form for playback later and at another speed compared to when it was stored.

Using the method of resetting the playout buffer to a certain size causes problems if there are very long talkspurts, e.g. broadcast from one speaker to several listeners. Since the length of a talkspurt is not defined in the beginning of the talkspurt the size to reset to might be either too small or too large. If it is too small, underrun will occur and if it is too large, unnecessary delay is introduced, thus the problem persists.

The general problem with the currently known approaches is that they are static and inflexible. As a conclusion dynamic solutions are required.

SUMMARY OF THE INVENTION

The present invention deals with the problem of improving speech quality in systems where the sampling rate at a transmitting terminal differs from the playout rate of a receiving buffer at a receiving terminal. This is often the case in e.g. IP-telephony.

When sampling frequencies are not controlled, underrun or overrun might occur in the playout buffer at the receiving side, which causes audible artefacts in the speech signal. To avoid said overrun or underrun there is a need for dynamically keeping the playout buffer to an average size, i.e. controlling the fullness of the playout buffer.

One object of the present invention is thus to provide a method for reducing audio artefacts in a speech signal due to overrun or underrun in the playout buffer.

Another object of the invention is to dynamically control the fullness of the playout buffer as not to introduce extra delay.

The above mentioned objects are achieved by means of dynamic sample rate conversion of speech frames, i.e. converting speech frames comprising N samples to instead comprise either N+1 or N-1 samples. More specifically the invention works on an LPC-residual of the speech frame and by adding or removing a sample in the LPC-residual, a sample rate conversion will be achieved. The LPC-residual is the output from an LPC-filter, which removes the short-term correlation from the speech signal. The LPC-filter is a linear predictive coding filter where each sample is predicted as a linear combination of previous samples.

By using the proposed sample rate conversion method, the playout buffer, of e.g. an IP-telephony terminal, can be continuously controlled with only small audio artefacts. Since the method works on a sample-by-sample basis the playout buffer

can be kept to a minimum and hence no extra delay is introduced. The solution also has very low complexity, especially when the LPC-residual already is available, which is the case in e.g. a speech decoder.

- 5 The term "comprises/comprising" when used in this specification is taken to specify the presence of stated features, integers, steps or components but does not preclude the presence or addition of one or more other features, integers, steps, components or groups thereof.
- 10 Although the invention has been summarised above, the method and arrangement according to the appended independent claims 1 and 23 define the scope of the invention. Various embodiments are further defined in the dependent claims 2-12 and 24-44.

BRIEF DESCRIPTION OF THE DRAWINGS

- 15 Figure 1 shows a transmitter and a receiver to which the method of the invention can be applied.
- Figure 2 shows a speech signal in the time domain.
- Figure 3 shows an LPC-residual of a speech signal in the time domain.
- 20 Figure 4 illustrates four modules of the sample rate conversion method according to the invention.
- Figure 5A shows an analysis-by-synthesis speech encoder with LTP-filter.
- Figure 5B shows an analysis-by-synthesis speech encoder with
- 25 adaptive codebook.
- Figures 5C-5F show different implementations of the LPC-residual extraction depending on the realisation of the speech encoder.
- Figures 5G-5J show four ways of placing the sample rate
- 30 conversion within the feed back loop of the speech decoder.
- Figure 6 illustrates how to use information about pitch pulses to find samples with low energy.

Figure 7 illustrates LPC-history extension.

Figure 8 illustrates copying of the history of the LPC residual.

DETAILED DESCRIPTION

5 The present invention describes, referring to figure 1, a method for improving speech quality in a communication system comprising a first terminal unit TRX1 transmitting speech signals having a first sample frequency F_1 and a second terminal unit TRX2 receiving said speech signals, buffering
10 them in a playout buffer 100 with said first frequency F_1 and playing out from said playout buffer with a second frequency F_2 . When the buffering frequency F_1 is larger than the playout frequency F_2 , the playout buffer 100 will eventually be filled with samples and subsequent samples will have to be discarded.
15 When the buffering frequency F_1 is lower than the playout frequency F_2 , the playout buffer will run into starvation, i.e. it will no longer have any samples to play on the output. These two problems are called overrun and underrun respectively, and causes audible artefacts like popping and clicking sounds in
20 the speech signal.

The above problems with underrun and overrun are solved by using dynamic sample rate conversion based on modifying the LPC-residual of the speech signal and will be further described with reference to figures 2-8.

25 Figure 2 shows a typical segment of a speech signal in the time domain. This speech signal shows a short-term correlation, which corresponds to the vocal tract and a long-term correlation, which corresponds to the vocal cords. The short-term correlation can be predicted by using an LPC-filter and
30 the long-term correlation can be predicted by using an LTP-filter. LPC means linear predictive coding and LTP means long

term prediction. Linear in this case implies that the prediction is a linear combination of previous samples of the speech signal.

The LPC-filter is usually denoted:

$$5 \quad H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^n a_i z^{-i}}$$

By feeding a speech frame through the LPC-filter, $H(z)$, the LPC-residual is found. The LPC-residual, shown in figure 3, contains pitch pulses P generated by the vocal cords. The distance L between two pitch pulses P is called lag. The pitch
 10 pulses P are also predictable, and since they represent the long-term correlation of the speech signal they are predicted through an LTP-filter given by the distance L between the pitch pulses P and the gain b of a pitch pulse P . The LTP-filter is usually denoted:

$$15 \quad F(z) = b \cdot z^{-L}$$

When the LPC-residual is fed through the inverse of the LTP-filter $F(z)$ an LTP-residual is created. In the LTP-residual the long-term correlation in the LPC-residual is removed, giving the LTP-residual a noise-like appearance.

20 The solution according to the invention modifies the LPC-residual, shown in figure 3, on a sample-by-sample basis. That is, an LPC-residual block comprising N samples is converted to an LPC-residual block comprising either $N+1$ or $N-1$ samples. The LPC-residual contains less information and less energy compared
 25 to the speech signal but the pitch pulses P are still easy to locate. When modifying the LPC-residual, samples being close to a pitch pulse P should be avoided, because these samples

contain more information and thus have a large influence on the speech synthesis. The LTP-residual is not as suitable as the LPC-residual to use for modification since the pitch pulse positions P are no longer available. As a conclusion, the LPC-residual is better suited for modification both compared to the speech signal and the LTP-residual, since the pitch pulses P are easily located in the LPC-residual.

The proposed sample rate conversion consists of four modules, shown in figure 4:

- 1) A Sample Rate Controller (SRC) module 400 that calculates whether a sample should be added or removed;
- 2) LPC-Residual Extraction (LRE) modules 410 are used to obtain the LPC-residual r_{LPC} ;
- 3) Sample Rate Conversion Methods (RCM) modules 420 find the position where to add or remove samples and how to perform the insertion and deletion, i.e. converting the LPC residual block r_{LPC} comprising N samples to a modified LPC-residual block r'_{LPC} comprising $N+1$ or $N-1$ samples; and
- 4) A Speech Synthesiser Module (SSM) 430 to reproduce the speech.

The idea behind the invention is that it is possible to change the playout rate of the playout buffer 440 by removing or adding samples in the LPC-residual r_{LPC} .

The SRC module 400 decides whether samples should be added or removed in the LPC residual r_{LPC} . This is done on the basis of at least one of the following parameters; the sampling frequencies of the sending TRX1 and receiving terminal units

TRX2, information about the speech signal e.g. a voice activity detector signal, status of the playout buffer or an indicator of the beginning of a talkspurt. These inputs are named SRC Inputs in the figure. On the basis of a function of one or
5 several of these parameters the SRC 400 forms a decision on when to insert or remove a sample in the LPC residual r_{LPC} and optionally which RCM 420 to use. Since digital processing of speech signals usually is made on a frame-by-frame basis, the decision on when to remove or add samples basically is to
10 decide within which LPC-residual r_{LPC} frame the RCM 420 shall insert or remove a sample.

There are basically three methods of obtaining the LPC-residual r_{LPC} that is needed as input to the RCM's 420. The methods depend on the implementation of the speech encoder and will be
15 described with reference to figures 5A-5F. The LRE solution also directly influences the SSM solution, which will become apparent below.

In figure 5A is an analysis-by-synthesis speech encoder 500
20 with LTP-filter 540 shown. This is a hybrid encoder where the vocal tract is described with an LPC-filter 550 and the vocal cords is described with an LTP-filter 540, while the LTP-residual $\hat{r}_{LTP}(n)$ is waveform-compared with a set of more or less stochastic codebook vectors from the fixed codebook 530. The
25 input signal S is divided into frames 510 with a typical length of 10-30 ms. For each frame an LPC-filter 550 is calculated through an LPC-analysis 520 and the LPC-filter 550 is included in a closed loop to find the parameters of the LTP-filter 540. The speech decoder 580 is included in the encoder and consists
30 of the fixed codebook 530 which output $\hat{r}_{LTP}(n)$ is connected to the LTP-filter 540 which output $\hat{r}_{LPC}(n)$ is connected to the LPC-filter 550 generating an estimate $\hat{s}(n)$ of the original speech signal $s(n)$. Each estimated signal $\hat{s}(n)$ is compared with the original speech signal $s(n)$ and a difference signal $e(n)$ is

calculated. The difference signal $e(n)$ is then weighted 560 to calculate a perceptual weighted error measure $e_w(n)$. The set of parameters that gives the least perceptual weighted error measure $e_w(n)$ is transmitted to the receiving side 570.

5

As can be seen in figure 5C the LPC-residual $\hat{r}_{LPC}(n)$ is the output from the LTP-filter 540. The SRC/RCM modules 545 can thus be connected directly to that output and integrated into the speech encoder. The LRE consists of the fixed codebook 530 and the long-term predictor 540 and the SSM consists of an LPC-filter 550, thus the LRE-module and the SSM-module are natural parts of the speech decoder.

If the speech encoder, on the other hand, is an analysis-by-synthesis speech encoder where the LTP-filter 540 is exchanged to an adaptive codebook 590 as shown in figure 5B, the LPC-residual $\hat{r}_{LPC}(n)$ is the output from the sum of the adaptive and the fixed codebook 590 and 530. All other elements have the same function as in figure 5A showing the analysis-by-synthesis speech encoder with LTP-filter 500. As can be seen in figure 5D the LPC residual $\hat{r}_{LPC}(n)$ is the sum of the output from the adaptive and fixed codebook 590 and 530. The SRC/RCM modules 545 can thus again be connected directly to that output and integrated into the speech encoder as shown in figure 5D. The LRE consists of the adaptive and the fixed codebook 590 and 530 and the SSM consists of an LPC-filter 550, thus the LRE module and the SSM module are again natural parts of the speech decoder.

When the speech encoder has some sort of backward adaptation, it is not feasible to make alterations in the LPC-residual since this would affect the adaptation process in a detrimental way. In figure 5E is shown how in these cases the parameters $\hat{s}(n)$ from the LPC-filter 550 could be fed to an inverse LPC-filter 525 placed after the speech decoder. After the sample

rate conversion has been made in the SRC/RCM modules 545 an LPC-filtering 550 is performed to reproduce the speech signal. The LRE module consists of the inverse LPC-filter 525 and the SSM module consists of the LPC-filter 550.

- 5 In figure 5F is shown how it is possible to produce an LPC residual $\hat{r}_{LPC}(n)$ through a full LPC analysis. The output $\hat{s}(n)$ from the speech decoder is fed to both an LPC analysis block 520 and an LPC-inverse filter 525. After the sample rate conversion has been made in the SRC/RCM modules 545, an LPC
10 filtering 550 is performed to reproduce the speech signal. The LRE consists in this case of the LPC analysis 520 respective the LPC inverse filter 525 and the SSM module consists of the LPC filter 550. Performing an LPC analysis is considered to be well known to a person skilled in the art and is therefore not
15 discussed any further.

- Referring again to figure 4, assume that the SRC-module 400 has decided that a sample should be added or removed in the LPC residual r_{LPC} and that the LRE module 410 has produced an LPC residual r_{LPC} . The RCM-module 420 then has to find the exact
20 position in the LPC-residual r_{LPC} where to add or remove a sample and performing the adding respective removing. There are four different methods for the RCM-module 420 to find the insertion or deletion point.

- 25 The first and most primitive method arbitrarily removes or adds a sample whenever this becomes necessary. If the sample rate difference between the terminals is small this will only lead to minor artefacts since the adding or removing is performed very seldom.

- 30 By inserting or removing samples at positions where the energy in the LPC-residual is low the synthesis will be less affected. This is due to the fact that segments close to pitch pulses

will then be avoided. To find these segments of low energy either a sliding window method or a simpler block energy analysis can be used.

5 The second method, called the sliding window energy method, calculates a weighted energy value for each sample in the LPC-residual. This is done by multiplying k samples surrounding a sample with a window function of size k ($k \ll N$), where N equals the number of samples in the LPC-residual. Each sample is then squared and the sum of the resulting k values is calculated.

10 The window is shifted one position and the procedure is repeated. The position where to insert or remove samples is given by the sample with the lowest weighted energy value.

The third method, block energy analysis, is a simpler solution for finding the insertion or deletion point. The LPC-residual

15 is simply divided into blocks of equal length and an arbitrary sample is removed or added in the block with the lowest energy.

The fourth method, as illustrated in figure 6, uses knowledge about the position P of a pitch pulse, and the lag L between two pitch pulses. With knowledge about that, it is possible to

20 calculate a position P' having low energy and where it is therefore appropriate to add or remove a sample. The new position P' can be expressed as $P' = P + k \cdot L$ where the constant k is selected so that P' is selected to be somewhere in the middle between two pitch pulses, thus avoiding positions with

25 high energy. A typical value of k is in the range of 0.5 to 0.8.

When the RCM-module 420 has calculated the position where to add or remove a sample it must be determined how to perform the insertion or deletion. There are three methods of performing

30 such insertion or deletion depending on the type of LRE-module used.

In the first method either zeros are added or samples with small amplitudes are removed. This method can be used for all LRE solution described above, see figures 5C-5F. Notice that in figures 5C and 5D the SRC/RCM-modules are placed before the synthesis filter SSM, but after the feed back of the LPC residual to the LTP-filter 540 respective the adaptive codebook 590.

In the second method insertion is carried out by adding zeros and interpolating surrounding samples. Deletion is performed by removing samples and preferably smoothing surrounding samples. This method can also be used for all of the LRE solutions described above, see figures 5C-5F. Notice that in figure 5C and 5D the SRC/RCM-modules are placed before the synthesis filter SSM, but after the feed back of the LPC residual to the LTP-filter 540 respective the adaptive codebook 590.

In the third method the SRC/RCM-modules 545 are placed within the feedback loop of the speech decoder, see figures 5G-5J, instead of after the feedback loop as in the previous methods. Placing the SRC/RCM-modules within the feedback loop uses real LPC residual samples for the sample rate conversion, by changing the number of components in the LPC-residual. The implementation differs depending on whether it is an analysis-by-synthesis speech encoder with LTP filter shown in figure 5A or an analysis-by-synthesis speech encoder with adaptive codebook shown in figure 5B, that is used.

For the speech decoder with LTP filter, see figure 5A, the SRC/RCM-modules 545 can be placed within the feedback loop in two different ways, either within the LTP feedback loop as shown in figure 5G or in the output from the fixed codebook 530 as shown in figure 5H. For the speech decoder with adaptive codebook, see figure 5B, the SRC/RCM can also be placed in two different ways, i.e. either before, figure 5J, or after, figure

5I, the summation of the outputs from the adaptive and the fixed codebook.

The alterations on the LPC residual consists of removing or adding samples just as before but since the SRC/RCM-modules 545
5 are placed within the LTP feedback loop some modifications must be done. The extending or shortening of a segment can be done in three ways either at the respective ends of the segment or somewhere in the middle of the segment. Figure 7 shows the case where the LPC residual is extended by copying two overlapping
10 segments, segment 1 and segment 2, from the history of the LPC residual to create the longer LPC residual. The normal case when no insertion or deletion is needed would be to copy N samples. Shortening the LPC residual is achieved by copying two segments that has a gap between them instead of being
15 overlapped. As before, it is important that a pitch pulse is not doubled or removed since this would introduce perceptual artefacts. Hence, an analysis should be performed in order to evaluate where to add or remove segments. This analysis is preferably made by using the same methods as discussed above
20 regarding how to find the position where to add or remove a sample in the RCM-module.

For all implementations except when the SRC/RCM-modules 545 are placed between the fixed codebook 530 and the LTP filter 540 the history of the LPC residual also has to be modified. The
25 lag L will be increased or decreased for the specific part of the history where a sample is inserted or deleted. Thus the starting position of the segment that will be copied from the history of the LPC residual, Pointer 1 or Pointer 2 in figure 8, needs modification. If the segment to copy is newer, i.e.
30 the case of Pointer 1, there is no need to modify the starting position. If, however, the segment to copy is older, i.e. the case of Pointer 2, then the pointer should be increased or decreased depending on if a sample is inserted or deleted. This

has to be managed for subsequent sub-frames and frames as long as the modification is within the history of the LPC residual.

When the SRC/RCM-modules are placed before the summation of the outputs from the adaptive and the fixed codebook as shown in figure 5J the length of the fixed codebook also needs to be changed. This is done by adding a sample, preferably a zero sample, in the output from the fixed codebook or removing one of the components. The insertion and deletion in the fixed codebook should be synchronised with the insertion and deletion in the adaptive codebook.

The invention being thus described, it will be obvious that the same may be varied in many ways. Such variations are not to be regarded as a departure from the scope of the invention, and all such modifications as would be obvious to a person skilled in the art are intended to be included within the scope of the following claims.

CLAIMS

1. A method for improving speech quality, in a communication system comprising a first terminal unit (TRX1), which transmits speech signals having a first sampling frequency (F_1) and a
5 second terminal unit (TRX2), which receives said speech signals, and buffers them in a playout buffer with said first frequency (F_1) and plays them out with a second frequency (F_2)
said method
- characterised by**
- 10 performing a dynamic sample rate conversion of a speech frame comprising N samples on a sample by sample basis, said dynamic sample rate conversion comprising the steps of
- creating an LPC-residual, comprising N samples, derived from
15 said speech frame;
- calculating, for each speech frame, whether a sample should be either added or removed from said LPC-residual;
- 20 generating a modified LPC-residual comprising N-1 or N+1 samples, if said calculating so demands; and
- synthesising a speech signal from said modified LPC-residual.
- 25 2. The method of claim 1 **characterised** in that the creating step comprises performing an LPC-analysis of the speech frame in order to find LPC-parameters of said speech frame.
3. The method of claim 1 **characterised** in that the creating
30 step comprises using already existing LPC-parameters from a speech decoder.
4. The method of claim 1 **characterised** in that the creating step comprises using an existing LPC-residual from a decoder.

5. The method of any of the preceding claims **characterised** in that the calculating step comprises deciding whether a sample should be added or removed on basis of at least one of the following inputs;

- the sample frequencies of sending (TRX1) and receiving (TRX2) terminal units;
 - a voice activity detector signal;
 - status of the playout buffer; and
- 10 - an indicator of the beginning of a talkspurt

6. The method of any of the preceding claims **characterised** in that the generating step comprises

selecting the position where in the LPC residual to add or remove a sample; and

performing said adding respective removing of said sample.

7. The method of claim 6 further **characterised** by selecting said position arbitrarily.

8. The method of claim 6 further **characterised** in that said position is found by searching for a segment of the LPC-residual with low energy.

9. The method of claim 8 further **characterised** in that said segment of low energy is found by using a block energy analysis.

10. The method of claim 8 further **characterised** in that said segment of low energy is found by using a sliding window energy analysis.

11. The method of claim 6 further **characterised** in that said position is found by using knowledge about the position of a pitch pulse together with knowledge about a time difference

between said pitch pulse and the following pitch pulse to select the position where to add or remove a sample in the LPC-residual.

5 12. The method of claim 11 further **characterised** in that said pitch pulse is found by searching for positions in the LPC residual with high energy.

10 13. The method of claim 12 further **characterised** in that said positions with high energy are found by using a block energy analysis.

15 14. The method of claim 12 further **characterised** in that said positions with high energy are found by using a sliding window energy analysis.

15. The method of claim 6 further **characterised** in that said adding of a sample is done by adding a zero sample.

20 16. The method of claim 6 further **characterised** in that said adding of a sample is done by adding a zero sample and interpolating surrounding samples.

25 17. The method of claim 6 further **characterised** in that said removing of a sample is done by removing a sample from the LPC-residual.

18. The method of claim 6 further **characterised** in that said adding of a sample is done by adding a sample in the history of the LPC residual; and

30 increasing a lag pointer as long as the adding is within the LPC residual history.

19. The method of claim 6 further **characterised** in that said removing of a sample is done by removing a sample in the history of the LPC residual; and

5 decreasing a lag pointer as long as the removing is within the LPC residual history.

20. The method of claim 6 wherein the second terminal unit comprises an adaptive and a fixed codebook

the method further **characterised** in that said adding of a sample is done by

10 adding a sample in the output from the adaptive codebook;

extending the output from the fixed codebook; and

increasing a lag pointer as long as the adding is within the LPC residual history.

15 21. The method of claim 6 wherein the second terminal unit comprises an adaptive and a fixed codebook

the method further **characterised** in that said removing of a sample is done by

removing a sample in the output from the adaptive codebook;

shortening the output from the fixed codebook; and

20 decreasing a lag pointer as long as the removing is within the LPC residual history.

22. The method of claim 6 wherein the second terminal unit comprises a fixed codebook

25 the method further **characterised** in that said adding or removing of a sample is done by

adding or removing a sample in the output from the fixed codebook.

23. An apparatus for improving speech quality in a communication system comprising a first terminal unit (TRX1) transmitting speech signals and having a first sampling
5 frequency (F_1) and a second terminal unit (TRX2) buffering said speech signals in a playout buffer with said first frequency (F_1) and playing them out with a second frequency (F_2), said apparatus

10 **characterised by**

means for performing a dynamic sample rate conversion of a speech frame comprising N samples on a sample by sample basis, said dynamic sample rate conversion further characterised by

15

means for creating an LPC-residual, comprising N samples, derived from said speech frame;

20 means for calculating for each speech frame whether a sample should be added or removed from said LPC-residual;

means for generating a modified LPC-residual comprising of N-1 or N+1 samples, if said calculating so demands; and

25 means for synthesising a speech signal from said modified LPC-residual.

24. The apparatus of claim 23 wherein the means for creating is
30 **characterised** by further comprising means for performing an LPC-analysis of the speech frame to find the LPC-parameters of said speech frame.

25. The apparatus of claim 23 wherein the means for creating is
35 **characterised** by further comprising means for using existing LPC-parameters from a speech decoder.

26. The apparatus of claim 23 wherein the means for creating is **characterised** by further comprising means for using an existing LPC-residual from a decoder.

5 27. The apparatus of any of claims 23-26 wherein the means for calculating is **characterised** by further comprising means for deciding if a sample should be added or removed on the basis of a function of at least one of the following inputs:

- 10 - sample frequencies of sending and receiving terminal units;
- a voice activity detector signal;
- status of the playout buffer; and
- an indicator of the beginning of a talkspurt.

15 28. The apparatus of any of claims 23-27 wherein the means for generating is **characterised** by further comprising
means for selecting the position where to add or remove
samples; and
means for performing said adding and removing.

20

29. The apparatus of claim 28 wherein the means for selecting is further **characterised** by means for arbitrarily selecting said position where to add or remove samples.

25 30. The apparatus of claim 28 wherein the means for selecting is further **characterised** by means for searching for the segment of the LPC-residual with the lowest energy.

30 31. The apparatus of claim 30 wherein the means for searching is further **characterised** by means for performing a block energy analysis.

35 32. The apparatus of claim 30 wherein the means for searching is further **characterised** by means for performing a sliding window energy analysis.

33. The apparatus of claim 28 wherein the means for selecting is further **characterised** by means for using knowledge about the position of a pitch pulse together with knowledge about a time difference between said pitch pulse and the following pitch pulse to select the position where to add or remove a sample in the LPC-residual.

34. The apparatus of claim 33 wherein the means for using knowledge about pitch pulses is further **characterised** by means for finding the pitch pulses by searching for positions in the LPC residual with high energy.

35. The apparatus of claim 34 wherein the means for finding pitch pulses is further **characterised** by means for performing a block energy analysis.

36. The apparatus of claim 34 wherein the means for finding pitch pulses is further **characterised** by means for performing a sliding window energy analysis.

37. The apparatus of claim 28 wherein the means for performing adding or removing is further **characterised** by means for adding a zero sample.

38. The apparatus of claim 28 wherein the means for performing adding or removing is further **characterised** by means for removing a sample from the LPC-residual.

39. The apparatus of claim 28 wherein the means for performing adding or removing is further **characterised** by means for adding a zero sample and interpolating surrounding samples.

40. The apparatus of claim 28 wherein the means for performing adding or removing is further **characterised** by

means for adding a sample in the history of the LPC residual;
and

means for increasing a lag pointer as long as the adding is
within the LPC residual history.

- 5 41. The apparatus of claim 28 wherein the means for performing
adding or removing is further **characterised** by

means for removing a sample in the history of the LPC residual;
and

- 10 means for decreasing a lag pointer as long as the removing is
within the LPC residual history.

42. The apparatus of claim 28 wherein the second terminal unit
comprises an adaptive and a fixed codebook

the apparatus further **characterised** by

- 15 means for adding a sample in the output from the adaptive
codebook;

means for extending the output from the fixed codebook; and

means for increasing a lag pointer as long as the adding is
within the LPC residual history.

- 20 43. The apparatus of claim 28 wherein the second terminal unit
comprises an adaptive and a fixed codebook

the apparatus further **characterised** by

means for removing a sample in the output from the adaptive
codebook;

- 25 means for removing a sample in the output from the fixed
codebook; and

means for decreasing a lag pointer as long as the removing is within the LPC residual history.

44. The apparatus of claim 28 wherein the second terminal unit comprises a fixed codebook

5 the apparatus further **characterised** by

means for adding or removing a sample in the output from the fixed codebook.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
220

ABSTRACT

The present invention relates to methods for improving speech quality in e.g. an IP-telephony system. The invention reduces audio artefacts being due to overrun or underrun in a playout buffer caused by the sampling rates at a sending and receiving side not being at the same rate. The inventive solution modifies an LPC-residual on a sample-by-sample basis. The LPC-residual block comprising N samples is converted to a block comprising N+1 or N-1 samples. A sample rate controller 400 decides whether samples should be added to or removed from the LPC-residual. The exact position where to add respective remove samples is either chosen arbitrarily or found by searching for low energy segments in the LPC-residual. A speech synthesiser module 430 then reproduces the speech. By using the proposed sample rate conversion method the playout buffer 440 can be continuously controlled. Furthermore, since the method works on a sample-by-sample basis the buffer can be kept to a minimum and hence no extra delay is introduced.

20 (Publication figure: Figure 4)

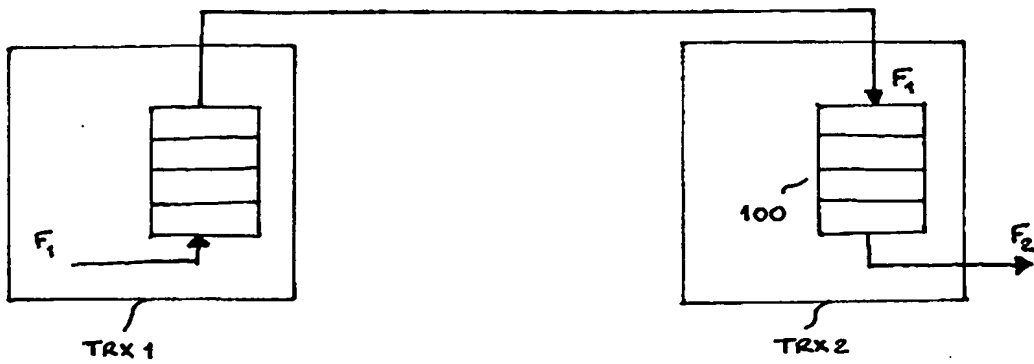


Figure 1.



Figure 2.

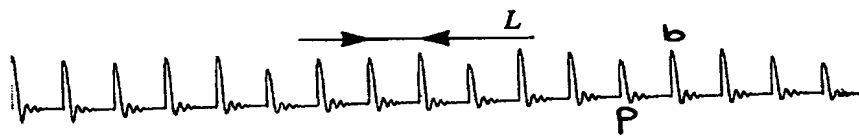


Figure 3.

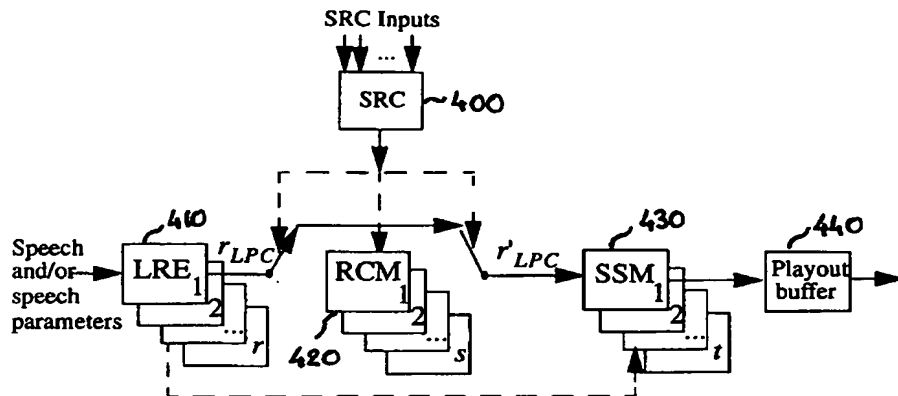


Figure 4.

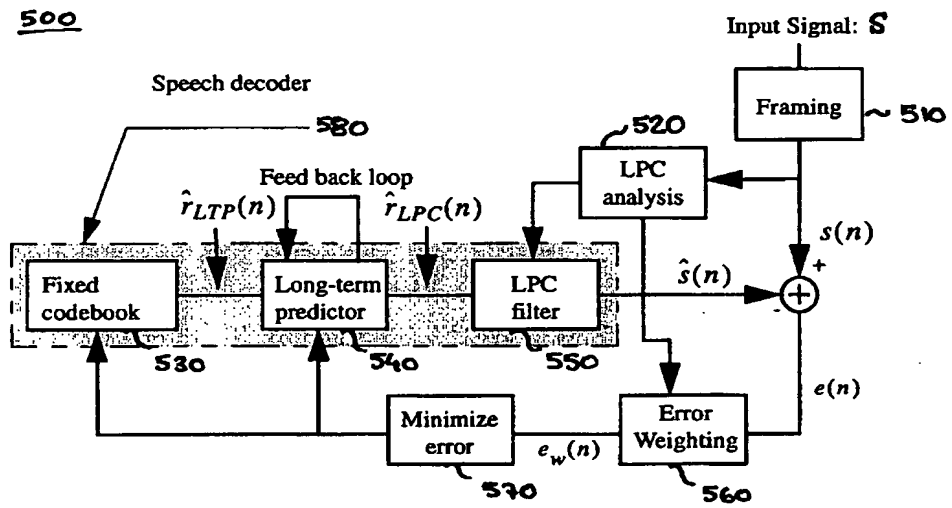
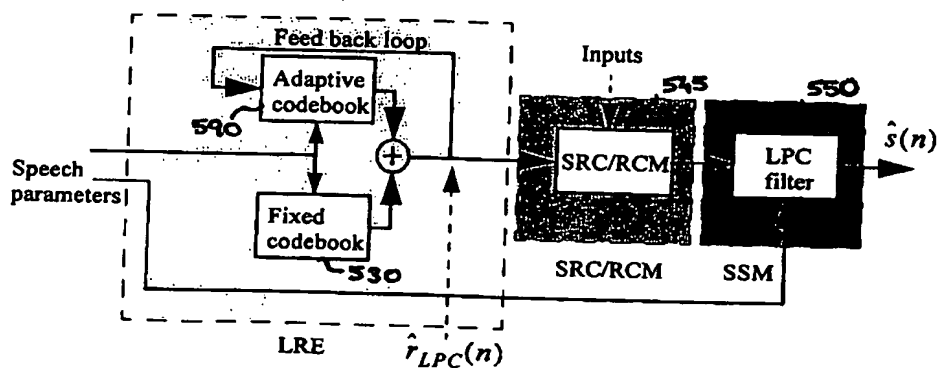
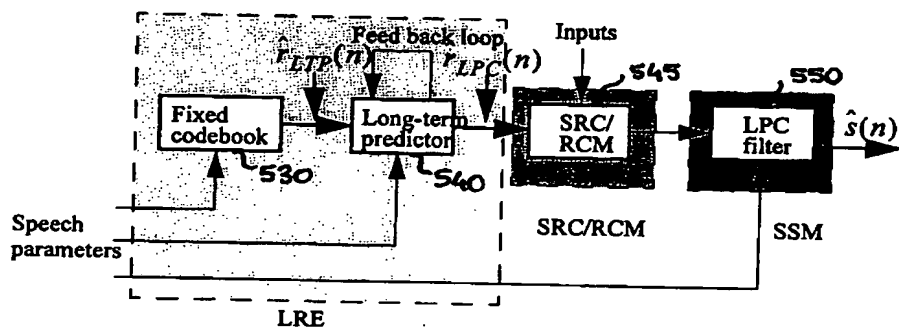
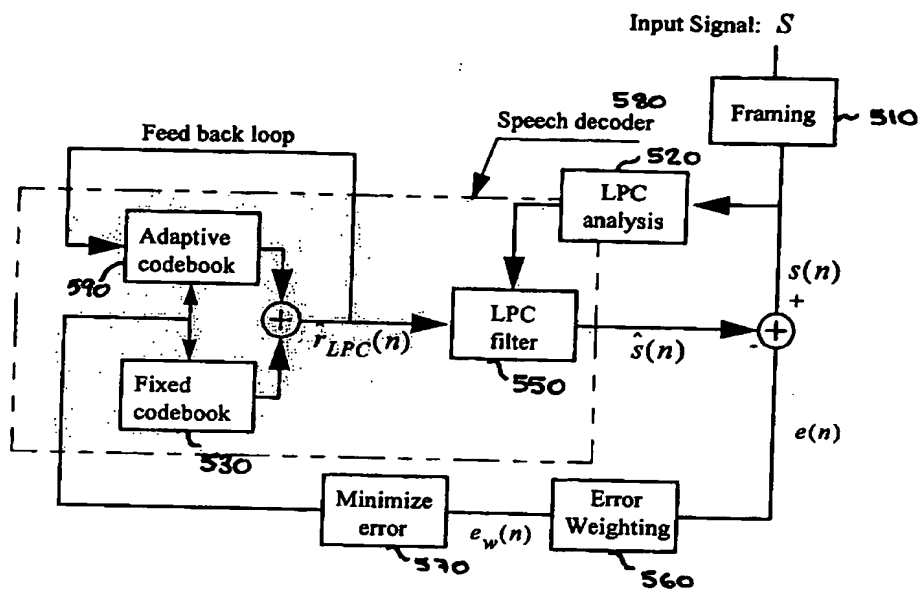


Figure 5A.



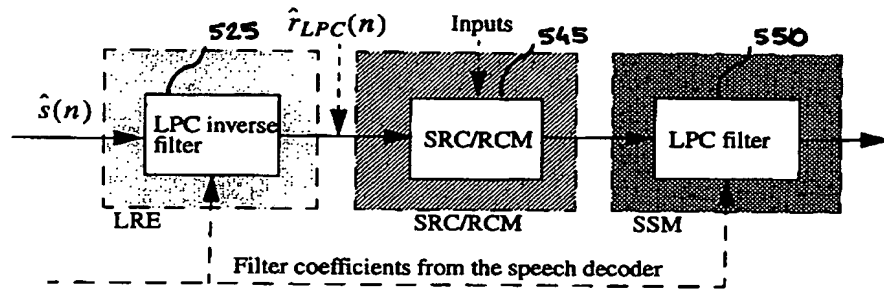


Figure 5E.

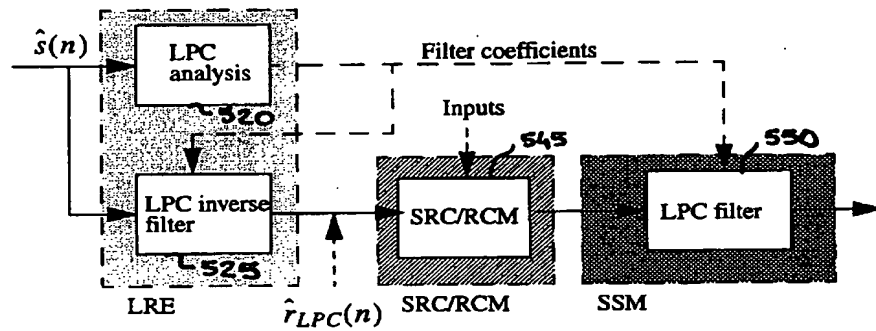


Figure 5F.

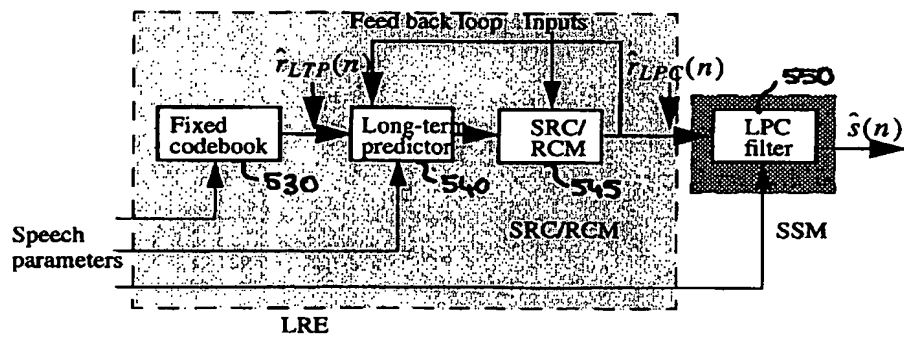


Figure 5G.

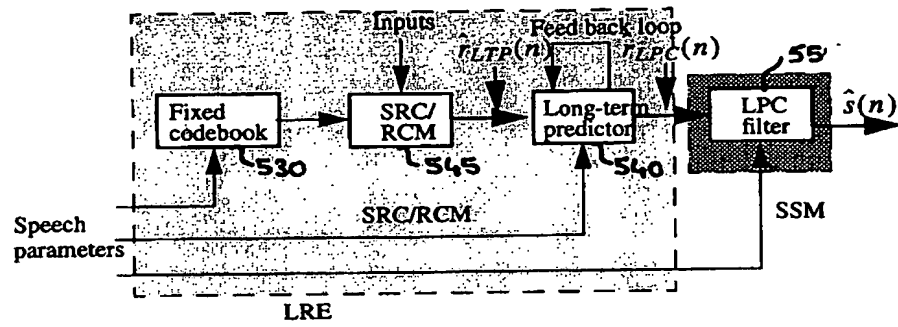


Figure 5H.

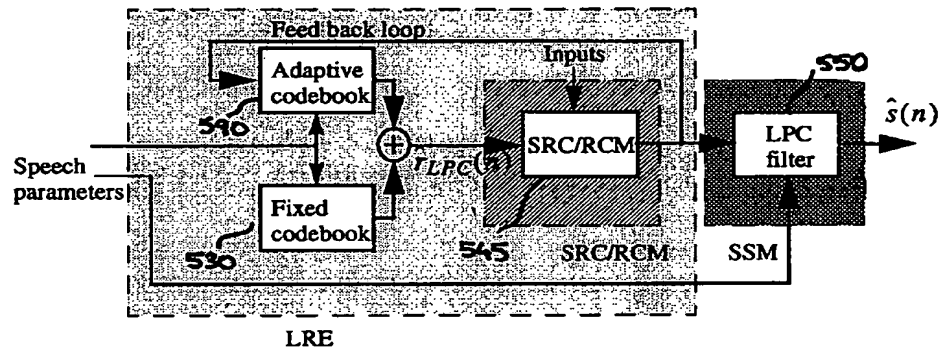


Figure 5I.

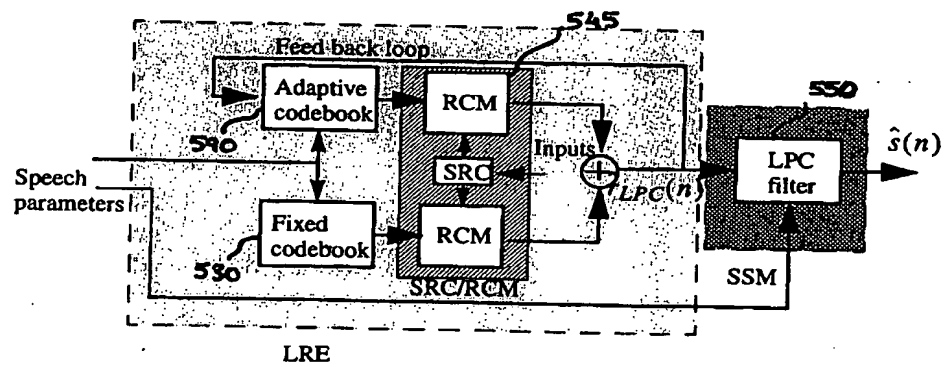


Figure 5J.

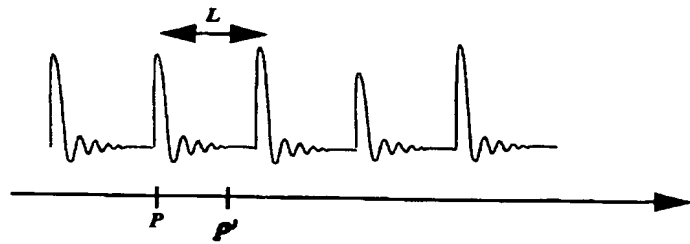


Figure 6.

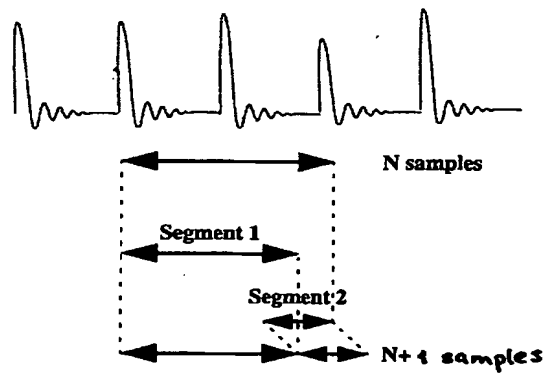


Figure 7.

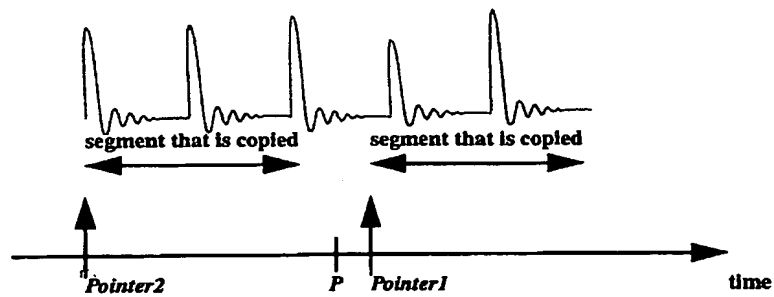


Figure 8.